PROMISE Annual Report 2011

http://www.promise-noe.eu/

PROMISE

promise

Participative Research labOratory for Multimedia and
Multilingual Information Systems Evaluation

With a population of over 500 million in its 27 states in which there are more than 80 indige-nous and many more immigrant languages, the citizens and companies of the EU demand information systems that allow them to interact with the culturally and politically diverse content that surrounds them in multiple media.

The technology and research behind multilingual and multimedia information systems are, today, in the position of intercepting these emerging trends but their design and develop-ment is becoming increasingly complex and needs proper means for ensuring that they meet the expected user requirements and provide the desired effectiveness.

Measuring is a key to scientific progress. This is particularly true for research and develop-ment concerning complex systems, whether natural or human-built. PROMISE will provide a virtual laboratory for conducting participative research and experimentation to carry out, ad-vance and bring automation into the evaluation and benchmarking of complex multilingual and multimedia information systems.

Summary of Activities

PROMISE is a Network of Excellence, started in September 2010 and scheduled until Au-gust 2013, with 10 academic and industrial partners, who are leaders in the field. PROMISE considers experimental evaluation – both laboratory and interactive – a key means for sup-porting and fostering the development of multilingual and multimedia information systems which are more adherent to the new user needs in order to ensure that they meet the ex-pected user requirements, provide the desired effectiveness and efficiency, guarantee the required robustness and reliability, and operate with the necessary scalability.

PROMISE aims at advancing the experimental evaluation of complex multimedia and multi-lingual information systems in order to support individuals, commercial entities, and com-munities who design, develop, employ, and improve such complex systems. The overall goal of PROMISE is to deliver a unified environment where to collect data, knowledge, tools, methodologies, and bringing together the user communities, which are involved in the ex-perimental evaluation.

Important Work Areas
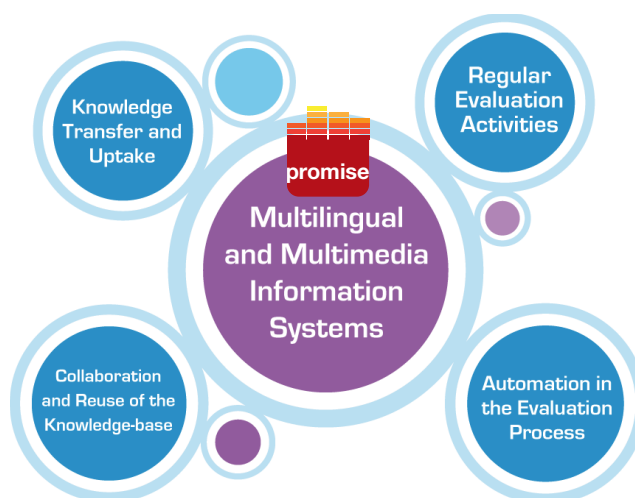
To achieve its goals, PROMISE pursues the following main activites:

- **Foster the adoption of regular and thorough experimental evaluation activities**: PROMISE carries on the successful and renowned CLEF (*Cross-Language Evaluation*

Forum) evaluation campaigns further pushing the evaluation exercises to tackle realistic tasks, use cases, and data sets;

- **Bring automation into the experimental evaluation process**: PROMISE proposes methods and provides a software infrastructure to create larger experimental collections; increase the number and size of the experiments conducted; and develop distributed, asynchronous, and loosely-coupled evaluation protocols, moving experimental evaluation from handicraft process to a mostly automatic one.
- **Promote collaboration and re-use over the acquired knowledge-base**: PROMISE will curate, preserve, and allow for the enrichment of the collected experimental data; it will provide the means for an easy comparison with and a meaningful interpretation and visualisation of the experimental results; and it will facilitate the discussion and collaboration among all the interested stakeholders.
- **Stimulate knowledge transfer and uptake**: PROMISE disseminates know-how, tools, and best practices about multilingual and multimedia information systems; facilitate uptake and participation by commercial entities and industries; and gives rise to multidisciplinary competencies and expertises.

CLEF 2011 – Conference on Multilingual and Multimedia Information Access Evaluation

Since 2000 the Cross-Language Evaluation Forum (CLEF) has played a successful role in stimulating research and promoting evaluation in a wide range of key areas in the information retrieval domain, becoming well-known in the international IR community. The results had traditionally been presented and discussed at annual workshops in conjunction with the European Conference for Digital Libraries.

As of 2010, CLEF has become and independent event whose aim is to continue with the benchmarking evaluation experiments, but also to provide more space for analysis, reflection and discussion. It is organised as a peer-reviewed conference interleaved by a series of laboratories and workshops. The conference aimed at advancing research into the evaluation of complex multimodal and multilingual information systems. The laboratories continue and expand on the traditional CLEF track philosophy. Two different forms of labs are offered: benchmarking activities proposing evaluation tasks and comparative performance analyses, and workshop-style labs that explore issues of information access evaluation and related fields.

CLEF 2011 was held at the University of Amsterdam, The Netherlands, September 19–22, 2011. Approximately 170 researchers from all around the world participated in the event.

CLEF 2011 featured six benchmarking activities: ImageCLEF, PAN, CLEF-IP, QA4MRE, Log-CLEF and new this year, Musi-CLEF. In parallel, there was the CHiC 2011 workshop dedicated to the evaluation of information access technologies in the setting of cultural heritage.


http://www.clef2011.org/

Four community sessions have been proposed to participants: the European Commission organized a session on "Intelligent Information Management - FP7 ICT Work programme 2011 - 2012, Strategic Objective 4.4. - General information and conditions of Call 8"; the CHORUS+ project discussed on "Information Retrieval from Scientific Multimedia Data"; PROMISE demoed its initial evaluation infrastructure prototype in the session "The Promise Evaluation Infrastructure, where Visual Analytics meets Information Retrieval"; finally, a ses-

sion on "Other Evaluation Initiatives" has been organized where the other major evaluation activities reported their recent outcomes.

The Initial Prototype of the Evaluation Infrastructure

The first prototype of the PROMISE evaluation infrastructure has been set up for carrying out the PROMISE use cases performed during CLEF 2011, namely "Search for Innovation", i.e. the CLEF-IP lab, and "Visual Clinical Decision Support", i.e. the ImageCLEF lab.
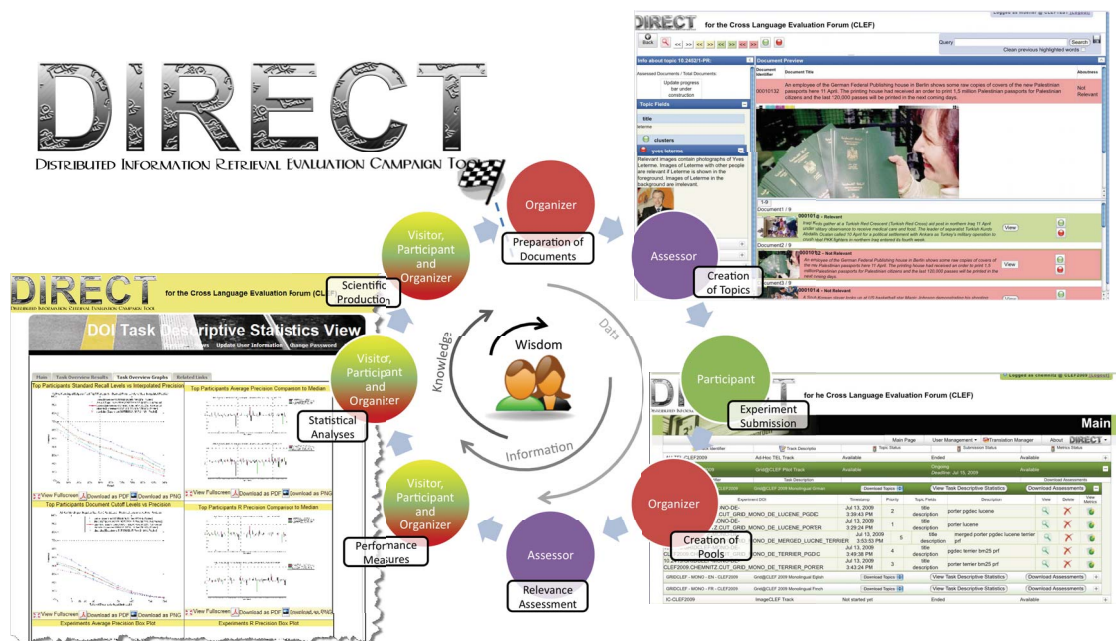
The first prototype allowed us to manage the submission of runs, computation of performance measures, distribution of the experimental results to the participants, and provision of on-line access to those results.

The overall goal of this initial prototype of the evaluation infrastructure is to actually test it in the context of the live CLEF 2011 Labs with actual multilingual and multimedia datasets and with real users in order to gather information, additional requirements, and feedback which serves for the release of the next prototype in the course of 2012.
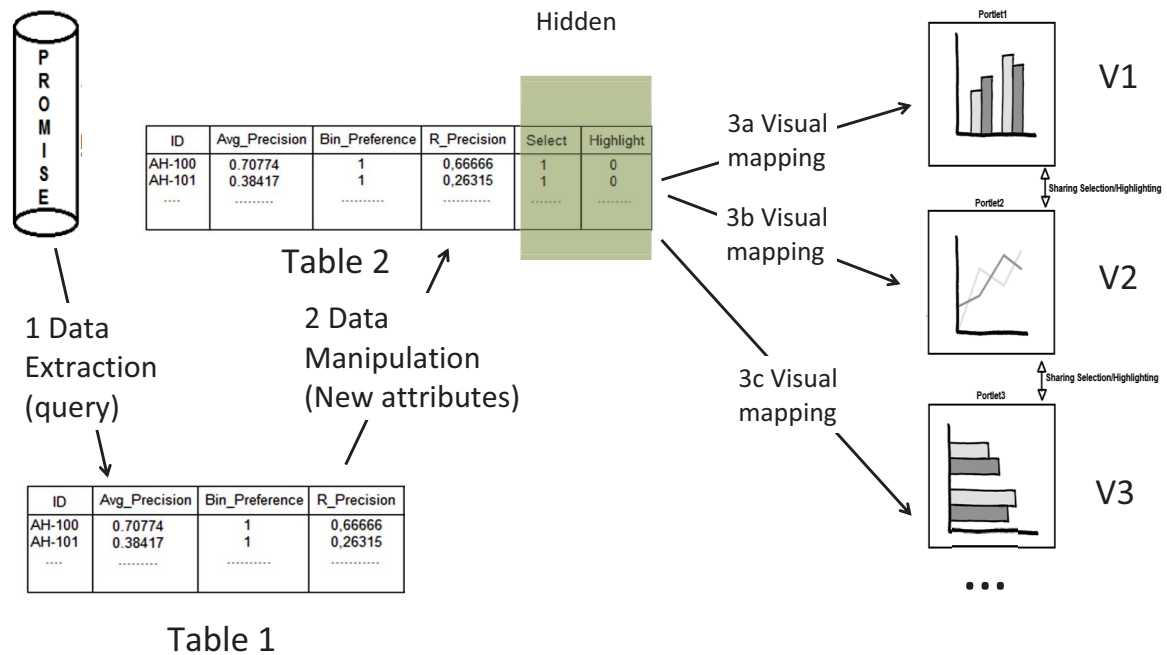
This initial prototype served also the purpose of early testing the innovative visual analytics environment which PROMISE is investigating and developing to provide advanced analysis and interpretation functionalities on the experimental data coming from information retrieval evaluation.

The most general situation is the one in which the system presents the user with multiple visualizations, each of them working on the same set of data. Visualizations are synchronized using two main interaction mechanisms: selection (it is just a way to focus the attention on a subset of data) and the highlighting (it allows for highlighting a part of the displayed data).

In order to produce a visualization, three main steps are, in principle, needed:

1) data extraction from the PROMISE infrastructure.

2) data manipulation, i.e., deriving new attributes, applying some aggregation operations , applying some analytical algorithms, etc. During such a process the system adds some hidden attributes to the data, in order to support the selection and the highlighting mechanisms.

3) Mapping the data obtained from step two on one or more visualizations.

The functioning of the visual analytics environment is shown in the figure below.

ID	Avg_Precision	Bin_Preference	R_Precision	Select	Highlight
AH-100	0.70774	1	0,66666	1	0
AH-101	0.38417	1	0,26315	1	0
....	....	....	....	....	....

Hidden

Table 2

3a Visual mapping

3b Visual mapping

3c Visual mapping

V1

V2

V3

•••

1 Data Extraction (query)

2 Data Manipulation (New attributes)

ID	Avg_Precision	Bin_Preference	R_Precision
AH-100	0.70774	1	0,66666
AH-101	0.38417	1	0,26315
....	....	....	....

Table 1

Use Case based Evaluation Activities

PROMISE proposes a use case based approach to evaluation. If the system with its various technological features is evaluated with respect to the designers' hypotheses of how the system should be utilised, the evaluation can proceed to validate or disprove those hypotheses without being distracted by confounding factors.

Use cases may be put together on very various levels of ambition, competence, and insight, but once formulated, interaction specialists can debate and test the validity of the use case; information system specialists can set parameters for system benchmarking, based on crucial characteristics of the use case; and industrial and commercial stakeholders can use a validated use case to build and design their systems with benchmarked system components for their purposes, once they find it conforms to their business case.

A use case is intended to capture all the ways a system is used by its environment, to describe all the services it offers and all the behaviour of the system and the actors engage in, for some specific purpose. The use case is a tool for developing a system, and user actions as formalised in the use case — most often using UML, the Unified Modeling Language — are mapped onto system components and system development objects for the purposes of system development and evaluation.

Scenarios, which often are the inspiration for use cases, are not use cases but instances of them: often several scenarios are necessary to track the various paths through a given use case for a system. A scenario describes the actions of a user during the course of an interaction.

A scenario describes the actions of a user during the course of an interaction. For instance, one scenario based on the use case "search for image of friend" in a description of an image search engine could be a story of Pyramus entering his friend Thisbe's name in the query field of the interface to find an oil painting by J W Waterhouse[1] of her.

[1] Image of Thisbe from `http://www.jwwaterhouse.com`.

PROMISE has as its starting point defined three use case domains:

- **Unlocking culture**: deals with information access to cultural heritage material held in large-scale digital libraries comprising libraries, archives, museum, and audio-visual archives.
- **Search for innovation**: deals with patent search and its peculiar requirements to seek out standardized method and framework for evaluating different tools for the IP.
- **Visual clinical decision support**: deals with visual information connected with text in the radiology domain in order to provide retrieval and access mechanisms able to jointly exploit textual and visual features.

Each of these use case domains can accommodate numerous use cases, and PROMISE has chosen to elaborate one sample in each domain. During 2011 they have been carried out in several activities at CLEF 2011 and, in particular, the CHiC 2011 workshop for the "Unlocking culture" use case, the CLEP-IP lab for the "Search for innovation use case", and the ImageCLEF lab for the "Visual clinical decision support".

User Involvement, Promotion and Awareness

PROMISE is constantly looking for opportunities to enlarge its community, reach close and neighbour communities, and target other relevant project and industrial stakeholders.

EuroVA 2011

EuroVA 2011 has been held in Bergen, Norway, on May 31st, 2011. It was the second international Eurovis workshop on visual analytics held in Europe and aims at continuing the success of the previous edition, held in Bordeaux, France, on June 8th, 2010. The goal of the workshop was to promote and advance the combination and integration of visualization and analytics methods for the purpose of problem solving in a variety of application domains including engineering, business, public policy, medicine, security, etc.

EuroVA 2011 represented a first opportunity to make aware the visual analytics community of the compelling issues that the experimental evaluation of multilingual information access system has to face when it comes to representing, interpreting, interacting, and exploring its experimental results. The event has been attended by more than 100 participants.

DESIRE Workshop at CIKM 2011

The *Data infrastructurEs for Supporting Information Retrieval Evaluation* (DESIRE) has been organized on 28 October 2011, co-located with the 20th ACM Conference on Information and Knowledge Management (CIKM 2011), in Glasgow, UK.

The main objective of the workshop wass to gather together experts from three areas – information retrieval, databases, and knowledge management – to encourage them to recognise the urgency of addressing the problem of envisaging and designing useful infrastructures able to coherently manage pertinent data collections and sources of information deriving from experimental evaluation, and so take concrete steps towards developing them in an integrated and coherent way.
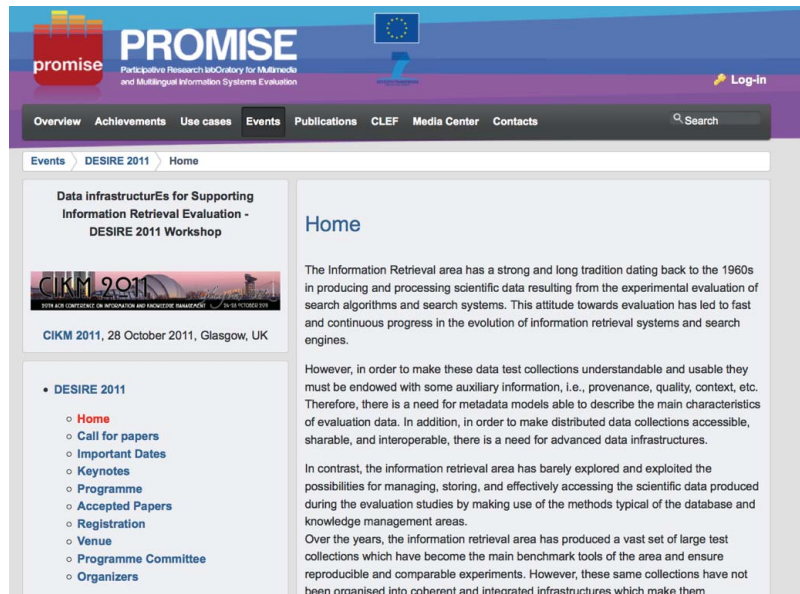
PatOlympics 2011

The PatOlympics Protoype Evaluation Exercise is ab innovative and very "athletic" competition designed to complement laboratory-style evaluation campaigns such as TREC-CHEM and CLEF-IP and to foster the interaction between creators of retrieval systems and information professionals.

The 2nd PatOlympics took place on 7 June 2011 in the morning and provided an interactive evaluation platform where prototypes of state-of-the-art patent retrieval systems have been tested and driven to their limits by patent retrieval professionals.

Participants were able to interact with the experts, demonstrate the power and flexibility of their prototype and compete with other participants for the medals of the PatOlympics.

PROMISE and Other Evaluation Initiatives

During CLEF 2011, the community session "Other Evaluation Initiatives" has been organized, where the following evaluation initiatives were represented:

- TREC, Ellen Voorhees, NIST, USA
- NTCIR, Hideo Joho, NII, Japan
- INEX, Jaap Kamps, University of Amsterdam, The Netherlands
- MediaEval, Gareth Jones, Dublin City University, Ireland

PROMISE and Other European Projects

PROMISE started collaborations and contacts with several European projects:

- **META-NET** (Technologies for the Multilingual European Information Society) is a Network of Excellence under FP7 dedicated to the technological foundations of the European multilingual information society.
- The **DC-NET** project (contract no. 246618) aims to develop and to strengthen the coordination of the public research programmes among the European countries, in the sector of the digital cultural heritage.
- The **Khresmoi** Integrated Project (contract no. 257528) aims at developing tools for medical information retrieval. The project is based on two use cases for visual and for textual and semantic information access.
- **CHORUS+** is a Coordination Action (contract no. 249008) that aims at creating the conditions of mutual information and cross fertilisation between the FP7 projects and the recently launched National Initiatives all over Europe together with Initiatives external to Europe.
- **CULTURA** is a STREP (contract no. 269973) that aims at delivering innovative adaptive services and an interactive user environment which dynamically tailors the investigation, comprehension and enrichment of digital humanities artefacts and collections.

Future Work

PROMISE will continue the exploration of use cases as an evaluation methodology and will produce a revised prototype of its evaluation infrastructure in order to support the CLEF 2012 evaluation activities.

Two main dissemination and knowledge transfer events will be organized as described in the following.

PROMISE Winter School 2012

The first PROMISE Winter school on "Information Retrieval meets Information Visualization" is being organized in Zinal, Valais, Switzerland, on 23 - 27 January 2012. The aim of the Promise Winter School on information retrieval and information visualization is to give participants a grounding in the core topics that constitute the multidisciplinary area of Multilingual Information Retrieval. The school is a week-long event consisting of guest lectures from invited speakers who are recognized experts in the field. The School is intended for PhD students, Masters students or senior researchers such as post-doctoral researchers form the fields of information visualization and information retrieval and related fields.

CLEF 2012

Work has already started for organizing to the CLEF 2012 conference which will be held in Rome, Italy, on 17-20 September 2012. CLEF 2012 will feature the following labs:

- **CHiC**: a last-year workshop turning into a benchmarking activity for the cultural heritage domain based on Europeana collections. It embodies the "Unlocking culture" use case.

- **CLEF-IP**: a benchmarking activity on intellectual property. It embodies the "Search for innovation" use case

- **ImageCLEF**: a benchmarking activity on image retrieval. It embodies the "Visual clinical decision support" use case.

- **INEX**: the well-known Initiative for Evaluation of XML retrieval joins efforts with CLEF to target new synergies between multilingual, multimodal and semi-structured information access.

- **PAN**: a benchmarking activity on plagiarism detection.

- **QA4MRE**: a benchmarking activity on the evaluation of Machine Reading systems through Question Answering and Reading Comprehension Test.

- **RepLab**: a benchmarking activity on microblog data for online reputation management.

A new workshop will be held:

- **eHealth**: a workshop on new evaluation issues in the health domain, related to the Louhi series of workshops on NLP in Health Informatics.

Further Information

Further information about PROMISE, its activities, and related events can be found at:
- CLEF 2011 home page
  http://www.clef2011.org/
- CLEF 2012 home page
  http://www.clef2012.org/
- CLEF General home page
  http://www.clef-campaign.org/

- Demos of the initial prototype of the PROMISE evaluation infrastructure
  http://www.youtube.com/watch?v=fDsXDCUPkiM
  http://www.youtube.com/watch?v=U73Qa8MAwYI
- PROMISE Winter School 2012
  http://www.promise-noe.eu/events/winter-school-2012/
- EuroVA 2011 Workshop
  http://www.eurova.org/previous-events/eurova-2011
- DESIRE Workshop at CIKM 2011
  http://www.promise-noe.eu/events/desire-2011/home

PROMISE Consortium

- University of Padua, Italy (coordinator)
  http://www.unipd.it/

- Swedish Institute of Computer Science (SICS), Sweden
  http://www.sics.se/

- University of Amsterdam, The Netherlands
  http://www.uva.nl/

- Sapienza University of Rome, Italy
  http://www.uniroma1.it/

- University of Applied Sciences Western Switzerland, Switzerland
  http://www.hes-so.ch/

- Zurich University of Applied Sciences, Switzerland
  http://www.zhaw.ch/

- Humboldt-Universität zu Berlin, Germany;
  http://www.hu-berlin.de/

- Evaluations and Language resources Distribution Agency (ELDA), France
  http://www.elda.org/

- Centre for the Evaluation of Language Communication Technologies (CELCT), Italy
  http://www.celct.it/

- Techinical University of Vienna, Austria
  http://www.tuwien.ac.at/

Contact Person

Dr. Nicola Ferro
Department of Information Engineering University of Padua
Via Gradenigo 6/b
35131 Padova, Italy
e-mail: ferro@dei.unipd.it
Tel: +39 049 827 7939
Fax: +39 049 827 7799